Even though every cell in the human body contains (almost) the same DNA sequence, they form different tissues and are characterised by structural and functional discrepancies, as an example, neurons are different to immune cells and other types of cells. This phenomenon arises from the fact that in cells of different types, different genes are turned on and, as a result, different mRNAs and proteins are produced in these cells.

Today, scientists are able to investigate individual cells to gain deeper insights into the diverse cell populations that form an organism. Single-cell RNA sequencing is now commonly used to illustrate the extent of heterogeneity between cells and to identify new subpopulations of cells. In order to learn more about the role of different cell types in regulating the function of a tissue or an organ, several computational methods to solve the problem of assigning types to cells in single-cell data were proposed.

While single-cell RNA data lacks information on cell position within the tissue, recently developed spatial transcriptomics (ST) technology enables to retain this information. Nonetheless, ST consists of RNA-sequencing measurements in multiple, distinct spots containing an unknown number of mixed cells. It is, therefore, necessary, to decompose the hidden cell-type mixture within each spot.

The aforementioned problems: 1) assigning types in single-cell data and 2) decomposing ST mixtures are demanding, of high importance and currently of interest to a vast number of scientists. In this study we will propose a novel computational method that will simultaneously perform the two tasks 1) and 2). To approach this, we will use probabilistic modelling and statistical bayesian inference methods. The model will be efficiently implemented and validated in Python.

In our opinion, the simultaneous performance of tasks 1) and 2) will permit us to gain an even higher resolution, and increased statistical confidence in comparison to the approach of performing them separately. As a result, we will contribute to solving the two problems 1) and 2), and by doing so to the understanding of heterogeneous tissues' functionality.

In addition, to demonstrate the model's usefulness we aim to analyse data on two types of tissues: the cortex of the mouse brain and a cancer affected breast tissue (to investigate the interactions of the malignant tumors with healthy cells in their neighbourhood).

In summary, we will deliver an innovative tool for the simultaneous cell type annotation in single-cell data and spatial transcriptomics cell-type-mixtures decomposition. The project will contribute to the understanding of functionality of heterogeneous tissues, in particular, the complex interplay of the tumour and its immune microenvironment in breast cancer. Crucially, the final and key result of our project shall be an easy to use, open source Python package, containing the model implementation, that can be used to analyse diverse tissues.